Africa RISING in Babati District, Tanzania

An impact evaluation concept note

[Draft, circulated for discussion; comments welcome]†

by

Ainsley Charles*

21 August 2013

Abstract. Using a prospective multi-arm randomized evaluation design and a new Tanzanian dataset collected for the purpose, a series of single-equation ANCOVA models will be used to estimate the causal effects of Africa RISING on several key agricultural and human development outcomes for the sub-population of agricultural smallholders targeted by, and responsive to, Africa RISING's outreach. The evaluation design described prejudices policy relevance over representativeness due to the way participants are recruited. The new evidence will contribute to the literature on the relative importance of training versus inputs, the effects of technological upgrading on farmer productivity and household welfare, the relative importance of credit, risk and skills constraints in smallholder agriculture, and to the broader literature on innovation adoption and diffusion in agriculture.

† I thank Carlo Azzarri, Mateete Bekunda, Beliyou Haile, Festo Ngulu, Joseph Rusike, David Spielman, and Africa RISING researchers and implementers for useful clarifications and discussions which informed this note.

* The author is Collaborator, Evaluation and Monitoring, Africa RISING East and Southern Africa Project, responsible for Malawi, Tanzania, and Zambia. He is based in Arusha.

Introduction

The Africa Research in Sustainable Intensification for the Next Generation (Africa RISING or AR) program is a five-year initiative funded by the United States Agency for International Development (USAID) (see Table 1). The program aims to "create opportunities for smallholder farm households to move out of hunger and poverty through sustainably intensified farming systems that improve food, nutrition, and income security, particularly for women and children, and conserve or enhance the natural resource base." (AR 2011.) The Consultative Group on International Agricultural Research (CGIAR) has joined with USAID to lead Africa RISING's three research-for-development projects. These include the International Institute of Tropical Agriculture (in West Africa and East and Southern Africa) the International Livestock Research Institute (in the Ethiopian Highlands), and the International Food Policy Research Institute (IFPRI) (all locations). IFPRI's key responsibility is to provide the services of monitoring, evaluation and impact assessment of the program. In keeping with this responsibility, this concept note provides the basis for IFPRI-led execution of a rigorous impact evaluation (IE) of a few elements of Africa RISING in one district in Tanzania.

According to a recent report that maps smallholder-focused international donor funding, as of 2009 a total of $6.7 billion was committed for smallholder support in Africa (FSG 2010). While other focus areas are access to finance or markets (34 percent) and infrastructure provision (22 percent), the largest share of smallholder-focused international donor funding in Africa was set to be the *direct provision of inputs or training* (43 percent). This focus of development aid means that research on the effectiveness of the direct provision of inputs and training will have high policy relevance. However, rigorous evidence on the effectiveness of each of these interventions is scarce, and their relative effectiveness and / or complementarity has to date not been tested.[1]

Objective

The objective of this concept note is to provide a plan for an impact evaluation of specific components of the Africa RISING program as it is currently being implemented in Babati District in the Manyara region of Tanzania. It takes into account the operational constraint that data-collection (which can be used as the impact evaluation baseline) is scheduled to happen within the coming month, while future roll-out will only be specified later. It can thus be seen as an evaluation of a pilot project that could inform the future scale-up and rollout of the program. As this happens, other evaluation possibilities may be determined.

This evaluation design therefore works within the following given parameters:[2]

1. The Africa RISING program, now ending its second operational year, did not include a prospective evaluation design at program launch.
2. An individual-level randomization has recently (July 2013) been done, which needs to be used for evaluation purposes.
3. A baseline survey is planned for August-September 2013.

[1] There is, however, an ongoing impact evaluation in Sierra Leone by the Abdul Latif Jameel Poverty Action Lab or J-PAL (Annan et al., not dated); see 'Previous Research'.
[2] The design also assumes that there will be a budget for its implementation.

4. Full rollout and scaling is still not planned at this point; Africa RISING remains essentially a pilot project.

Table 1. Details of the Africa RISING program

Item		*Remarks*
Name		Africa Research in Sustainable Intensification for the Next Generation (RISING)
Objective		To create opportunities for smallholder farm households to move out of hunger and poverty through sustainably intensified farming systems that improve food, nutrition, and income security, particularly for women and children, and conserve or enhance the natural resource base.
Program type		Analysis of farming systems; demonstration of new agricultural technologies
Dates of implementation		2011 to 2016
Delivery mechanism		Research agencies (proposals, technical studies, implementation plans), in collaboration with village authorities and extensionists, on individual but interlocking 'work packages'
Outputs		Seed and fertilizer trials, land characterization surveys, etc.
Annual budget		$3.2 million (US) for East and Southern Africa (2012-2013)
Targeting Primary group Geographic Household / individual		 Agricultural smallholders Yes Yes
Targeting criteria Economic Environmental Social		 Yes Yes No

Source: Author's compilation

Background

The United Republic of Tanzania is an LDC in East Africa with a poverty incidence in 2007 of 33.6 percent on the mainland (IMF 2012). The economy has been strengthened in recent years by improved performance in, among other sectors, construction, mining and quarrying, and communications, but agriculture remains the major provider of employment and income in rural areas. Recent agricultural growth has been driven by increases in cultivated acreage, rather than in productivity, and the sector remains challenged by underdeveloped infrastructure, insufficient extension, and inadequate research (IMF 2012). In respect of agriculture, maize and rice represent the two main staple crops in Tanzania, and maize production doubled in the five years to 2007, from 2.6 million tonnes to 5.4 million tonnes (NBS 2010).

Babati District, with 60,000 households, is the second largest district (by population) in the Manyara region of northeastern Tanzania, where just over one in every four smallholder crop-growing

households used improved seeds in 2007, while one in five used fertilizer (NBS 2012). It comprises twenty-one *shehia*, or wards, including Bashnet, Dabil, and Dareda, home to the Africa RISING villages of Long, Sabilo, and Seloto, respectively. As in other sites of operation, AR in Tanzania comprises a collection of interventions, or work packages, aimed collectively at "providing pathways out of hunger and poverty through sustainable intensification" (AR 2011). Multiple research partners for each work package exist. Research team leaders include ARI-Hombolo, AVRDC, CIAT, CIMMYT, ICRAF, ICRISAT, IITA, ILRI, and Selian ARI.[3]

The program comprises seven work packages in Babati District and is managed by IITA (along with its sister program in the districts of Kiteto and Kongwa). The second work package (WP2) involves dissemination, using farmer field days, of improved crop varieties to farmers, including maize, pigeon pea, and common bean. These are specifically bred for high yield and drought and disease tolerance favored by agronomists and geneticists, and enhanced quality characteristics such as shorter maturity and improved texture favored by farmers. Ten varieties were planted in January-February 2013 at community demonstration sites (in the sub-villages of Baryomod, Dulaghang, and Gitenyamur in Sabilo, and Dactara-B and Semak-B in Seloto). These include four experimental hybrids produced by CIMMYT-Kenya (CKH08051, CKH10058, CKH10717, and CKH101795), three by private seed companies (MH616, SAH536, and SAH538), two by Selian-ARI (SH208 and SH308), and one pre-existing local variety, SC627 (to serve as a control). And, in an illustration of the integrative model employed by Africa RISING, these demonstration activities were combined with practical advice on application of local inorganic fertilizer (*Minjingu Mazao)*, the focus of the Selian-ARI-led fourth work package. The most recent field days were held in July 2013.

Work-Package-1 (WP1) is geared to identify biophysical constraints using agronomic and soil surveys and the approach of farming systems analysis; it is led by CIAT. ILRI leads the third work package, whose focus is on introducing improved fodder species into maize-based systems as a land management strategy, while the fifth, sixth, and seventh packages involve, respectively, assessment of maize and groundnut mycotoxin contamination (IITA-led), integration of postharvest nutrition technologies into maize-based farming (also IITA-led), and a vegetable oriented intervention (led by AVRDC).

Previous research

The World Bank's 2011 report "Impact Evaluations in Agriculture – An Assessment of the Evidence" identifies input technology (improved seeds and fertilizer use), as one of the four areas in impact evaluation in agriculture in which further rigorous counterfactual-based research is needed. There is, however, ongoing work that will generate results in the near future. In a project in Mozambique, Carter and Yang (not dated) are testing interactions between fertilizer subsidy vouchers and various savings incentives designed to help farmers to afford fertilizer. In a project with similar aims to the impact evaluation proposed here, Annan et al. (not dated) are testing the interactions between free or subsidized access to improved seeds and a training program. Due to a larger sample size, they are able

[3] AVRDC: The World Vegetable Center; CIAT: International Center for Tropical Agriculture; CIMMYT: International Maize and Wheat Improvement Center; ICRAF: World Agroforestry Center; ICRISAT: International Crops Research Institute for the Semi-Arid Tropics; IITA: International Institute of Agriculture; ILRI: International Livestock Research Institute; Selian-ARI: Selian Agricultural Research Institute.

to test against each other various subsidy levels (from zero percent (market price), to 50 percent and 100 percent (free provision)), and test each of their interactions with the training program.

In the area of knowledge dissemination and training, impact evaluations have largely focused on studying the effects of agricultural extension and of farmer field schools. None of these are experimental studies and, with the exception of Dercon et al., do not test different productivity-enhancing interventions against each other. Positive effects of extension on farm and / or household-level outcomes were shown by Benin et al. (2008) and Bolwig et al. (2009) for Uganda, Dercon et al. (2008) for Ethiopia, Owens et al. (2002) for Zimbabwe. On farmer field schools, the evidence is more mixed with Feder et al. (2004) finding no significant effects for an Indonesian project, while Godtland et al. (2004) find very tentative positive effects on yields in Peru.

Evaluation research questions

The primary aim of the proposed evaluation is to determine a) whether training on the use of advanced agricultural technologies on its own can boost adoption and ultimately improve agricultural yields and household welfare in Babati District, b) whether direct provision of these technologies can, by itself, achieve similar results, or c) whether there are complementarities such that only a package of the two is effective.

In particular, this means answering the following research questions:

1. Is the provision of ***inputs*** (fertilizer and / or improved seeds) associated with a direct effect on a series of key outcomes in treated farming households, as compared to untreated farming household?
2. Is the provision of ***training*** (on use of fertilizer and improved seeds, on recognition of disease ("maize doctor") associated with a direct effect on a series of key outcomes in treated farming households, as compared to untreated farming households?
3. Is the provision of ***inputs-plus-training*** associated with a direct effect on a series of other key outcomes in treated farming households, as compared to untreated farming households?
4. Which of the above three treatments has the largest effect on farm- and household-level outcomes?

From a theoretical point of view, this evaluation setup will also help answer the question whether the binding constraint to adoption of advanced technologies, higher productivity and improved human development outcomes among targeted Babati farmers is credit- and risk-based (see research question 1), skills-based (see research question 2), or both – such that relaxing one constraint only will not be sufficient (see research question 3).

Results chain and key outcome variables

Answering the research questions proposed above will require measuring the effects of the interventions all along the results chain, from knowledge to behavioral changes and ultimately farm-level outcomes and household welfare.

The proposed list of outcomes includes the following:

1. Knowledge on advanced technologies
    a. Awareness that new technology exists
    b. Perception that it is useful
    c. Knowledge of how to implement it
2. Behavioral changes: Input use
    a. Improved seeds purchased and planted
    b. Local inorganic fertilizer purchased and applied
    c. Use of other inputs (land, labor, capital, conventional seeds, organic fertilizer)
3. Agricultural outcomes
    a. Yield per acre
    b. Sales
    c. Income
4. Household outcomes
    a. Consumption
    b. Food security
    c. Employment
5. Human development outcomes
    a. Child outcomes
    b. Women of reproductive age outcomes
    c. Subjective welfare

Any indicators listed here that are not part of the current questionnaire draft could still be incorporated in the final version. The unit of analysis is the household for most of the key outcome variables with some additional individual-level outcomes, so agricultural and welfare measurements will be taken at these levels (see 'Data' below).

Design

The objective of the proposed evaluation is to determine a) whether training on the use of advanced agricultural technologies on its own can boost adoption and ultimately improve agricultural yields and household welfare, b) whether direct provision of these technologies can, by itself, achieve similar results, or c) whether there are complementarities such that only a package of the two is effective. Accordingly, the experimental setup will consist of three treatment groups and a control group. Group 1 will receive only inputs, Group 2 only training, and Group 3 both, while the control group will receive neither. With a proposed total sample size of 1,000 farming households and equal-sized groups (see 'Power calculations' below) this implies the setup depicted in Table 2.

For the purposes of this evaluation, the intervention protocols are:

- Inputs: improved seeds, local inorganic fertilizer (*Minjingu Mazao*), or both. Inputs will be provided in October 2013 in preparation for new season planting in November. Inputs will be provided for one time only at the beginning of the agricultural season at a rate of two 50-kg bags per acre.

- ➢ Training: either a full AR package that includes demonstration of improved maize hybrids, intercropping technologies, and the correct application of local inorganic fertilizer (WP2 and 4), 'maize doctor' and soil management training (WP1), mycotoxin contamination recognition training (WP5), and post-harvest management training (WP6), or a restricted package of training relevant to the inputs being provided. Training will be provided to farmers in October 2013, also in preparation for the new agricultural season, but may be repeated or sequenced during the year.

Table 2. AR in Babati: Experimental setup, by intervention status

		Treatment 2 (Training)	
		Yes	No
Treatment 1 (Inputs)	Yes	Group 3 (250)	Group 1 (250)
	No	Group 2 (250)	Control (250)

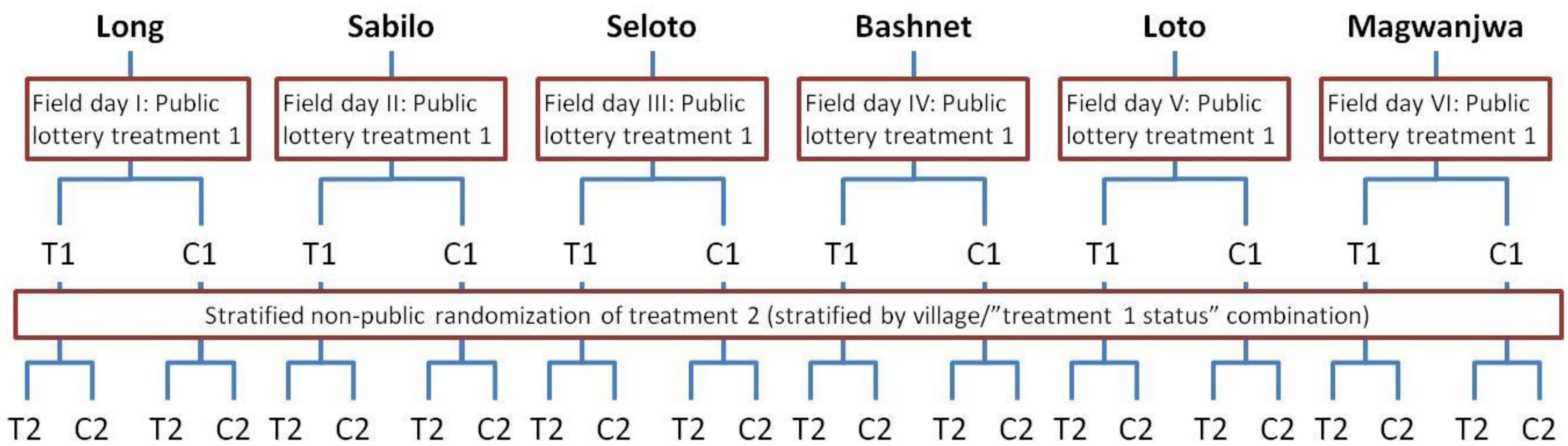
Source: Author's compilation

To achieve this design, I propose to expand the sample of an inputs voucher randomization that has already been carried out as well as add a second layer of randomization for the training intervention. This procedure is illustrated in Figure 1.

At the first stage, participant recruitment will be facilitated by mobilization at the village and sub-village level to attend a field day. Of those attending, half will be randomly selected, via public lottery, to receive inputs (local inorganic fertilizer and / or improved seeds) to begin the new agricultural year. The public lotteries have already been carried out in the three villages of Long, Sabilo and Seloto at field days in July, and, to ensure power (see 'Power Calculations' below), a second set of villages will be targeted during early September. As these villages do not have demonstration plots, farmers could be invited to the existing demonstration plots in Long, Sabilo and Seloto.

The mobilization, field day and lottery procedures in the new villages have to replicate those applied in the first round. This approach is required to make sure that all farmers in the sample are comparable in that they all could be mobilized to come to a field day (a self-selection that makes them different on average from a randomly selected farmer) and in that they have received the same field-day intervention. The three proposed additional villages in Babati District are Bashnet (same ward as Long), Loto (same ward as Seloto), and Magwanjwa (same ward as Sabilo). They are similar in terms of agro-ecology and farming systems to the existing three villages, and had for that reason been suggested by district-level extensionists when a different evaluation approach was being considered.

Figure 1: Two-stage randomization in 6 villages in Babati



*Note: T1 refers to the group randomized to receive treatment 1 (inputs), C1 to the group randomized not to receive it. The same applies for T2 and C2 with respect to treatment 2 (training).

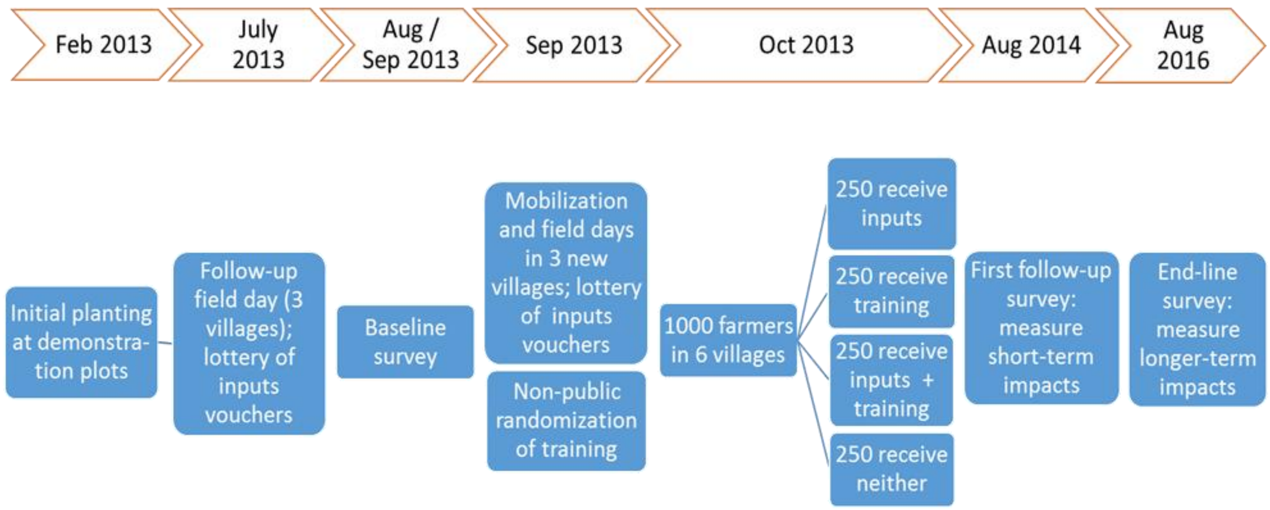
At the second stage, I propose a non-public randomization of the training intervention among the full sample of farmers who came to the field days in the six villages. This randomization will be stratified at the village-'input treatment status' level to assure balance: within each village, half of those who were randomly selected to receive the input voucher will receive the training, as will half of those who were randomized not to receive the input voucher. This randomization should be non-public even for the villages that have not yet had a field day to ensure a uniform methodology across the whole sample.

The chosen approach of using a sample of farmers who attended a field day means that the evaluation will not measure the average impact for all Tanzanian farmers, or for all farmers within Babati District. Instead, impacts are measured for farmers who would typically respond to the outreach activities of Africa RISING (and of programs with similar mobilization and sensitization modalities) and results are therefore highly policy-relevant.

Furthermore, as in the current setup the IE is integrated with the monitoring of the project within the responsibilities of IFPRI, I will be able to conduct cost-benefit analysis; the IE will provide a valid estimate of the causal impact while the monitoring data will contribute precise data on costs of the two different interventions at the project level. Therefore, I will be able to calculate costs-per-incremental-yield-increase to compare the cost-effectiveness of the two interventions, as well as to compare the cost-effectiveness of AR with other similar programs.

Figure 2 presents a timeline for the proposed impact evaluation. In February 2013, ten varieties of maize, along with beans and pigeon peas, were planted at community demonstration sites. At follow-up field days in three villages in July 2013, approximately 440 farmers attended, assessed and discussed pre-harvest realizations with scientists and project implementers, and independently ranked preferred seeds. In addition, public lotteries for input vouchers were held. The baseline survey is planned for August / September 2013. In September 2013, field days with the same program including the public randomization shall be held for farmers from an additional three villages. The full sample of farmers who attended the two sets of field days shall then be assigned to training treatment or control status in a stratified, non-public randomization as described above. Finally, in October 2013 (just before the next planting season), farmers shall receive their "treatments" as dictated by the randomization results.

Figure 2. Africa RISING in Babati District: Activity timeline for impact evaluation



Source: Author's compilation

This plan assumes the ideal case that funding can be assured for two follow-up surveys, the first one year after the baseline (August 2014) to capture short-term impacts early on in the results chain (knowledge, behavioral changes, initial effects on agricultural outcomes; see 'Results chain and key outcome variables'), the second three years after the baseline (August 2016) to measure effects that take longer to materialize, in particular the full set of agricultural outcomes, as well as household and human development outcomes. The Millennium Challenge Corporation has recently emphasized the importance of measuring both early and longer-term results in agricultural impact evaluations, especially from an accountability perspective; see Farley et al. (2012).

Method

For estimation of the impact of input and training provision, data collected will be used to compare mean outcomes of beneficiaries of different interventions (treatment groups) with each other as well as with mean outcomes of non-beneficiaries (control group). As previously described, the design includes three treatments that households will be assigned to randomly, which produces groups that are comparable along key dimensions, including observable and unobservable characteristics that may be predictive of outcomes of interest. For all outcomes the study will estimate intention-to-treat effects, recovered through a series of simple mean difference and analysis of covariance (ANCOVA) models.

The following specification will be initially used to recover the effect of Africa RISING's input / training intervention on household and individual outcomes:

$$Y_{it} = \alpha + \beta_1 I_{it} + \beta_2 T_{it} + \beta_3 (I * T)_{it} + \epsilon_{it} \qquad (1)$$

where it subscripts identify unit of observation i (household or individual, as warranted) and time period t, Y stands for outcome measures (previously identified), I and T are the treatment variables of

interest (inputs, training, or their combination) which are equal to 1 if unit i has received inputs (or training) and 0 otherwise, the α and βs are the parameters to be estimated, and ϵ represent independent and identically distributed disturbance terms. All impact estimates will be based on standard errors that have been adjusted to account for the likely lower variability of outcomes among individuals within the same village as compared to individuals of different villages.

To increase precision of the estimates, this basic model will be extended to include – as additional covariates – baseline values of the outcome measures, producing the ANCOVA model. The extended specification is given by:

$$Y_{it} = \alpha + \beta_1 I_{it} + \beta_2 T_{it} + \beta_3 (I * T)_{it} + \beta_4 Y_{it}^{BL} + \epsilon_{it} \qquad (2)$$

where BL superscript indicates that outcomes are measured at baseline.[4]

Power calculations

The table below shows sample size calculations that demonstrate what sample sizes are required in order to detect effects of various sizes. The indicator used for these power calculations is maize yield per acre (in kilograms), a central outcome of interest. To get the pre-intervention average maize yield and its standard deviation that are required for these computations, I use the National Bureau of Statistics 2007-08 Agricultural Sample Census, which sampled 53,000 rural agricultural households and fully enumerated 1,000 large farms on mainland Tanzania and in Zanzibar. Average maize yield in the long (*masika*) season for mainland Tanzania, Manyara region, and Babati district was 585.9 kg per acre (standard deviation: 512.5), 701 kg per acre (s.d.: 528.2) and 892.8 kg per acre (s.d.: 537.3), respectively. Using both regional (Manyara) and district (Babati) data as baseline values, sample sizes are calculated for effect sizes of 10, 15, and 20 percent increases in yield, as well as of one-fifth and one-half standard deviation increases in yield.[5]

Table 3 shows the total sample size required to detect such effects.[6] Row 1 represents the preferred model that uses only Babati data to approximate pre-intervention average and standard deviation of maize yield. Further specifications are power 0.8, significance criterion 0.05, no baseline data and equal-sized groups. To detect an increase in Babati maize yield to 982.1 kg per acre (that is, an increase of 10 percent) due to the inputs-or-training intervention, with power of 0.8, a sample size of 2,276 farmers is required. Detecting an achieved maize yield increase of 133.7 kg per acre (a 15 percent increase) requires 1,012 farmers, while detecting a 20 percent increase requires only 572 farmers.

In each of rows 2 to 5, one specification of the preferred model is changed. In row 2, mean and standard deviation derived from all of Manyara region are used, as opposed to just Babati district.[7] Due to a larger standard deviation, sample sizes required for each of the effect sizes are larger. Similarly, row 3 shows

4

5 For Manyara, I adjust the data by dropping observations with yield per acre above 2,350, the maximum in Babati, our district of interest. This was also given as the plausible national maximum in the enumerator guidelines accompanying the census. This means dropping two observations.

6 All power calculations use Stata's *sampsi* command (Stata 2009).

7 I include the Manyara data as a robustness check to account for the fact that the Babati averages and standard deviations are based on relatively fewer observations.

that increasing power to 0.9 increases the minimum required sample size. Finally, in rows 4 and 5 I account for the fact that there will be a baseline, which in theory saves on the number of observations required for a given effect size. I assume an inter-temporal coefficient of similarity of 0.04 (row 4) and 0.2 (row 5).The former is derived from the Ethiopian Rural Household Survey (ERHS; 2005, 2009) because the Tanzania sample censuses of 2002 and 2007 are not panel data, and the latter is an assumption for illustration purposes. For very low levels of inter-temporal correlation, the reduction in minimum sample size is irrelevant in practice (see row 4), while even for a slightly larger inter-temporal correlation the numbers do not change much (see row 5).

Table 3. Sample Size Calculations for AR Babati Inputs-or-Training Intervention

			Effect size				
			+10%	**+15%**	**+20%**	**+0.2sd**	**+0.5sd**
1.	**Preferred Model: Babati data, Power 0.8, Sign. 0.05, No BL**		2276	1012	572	1572	252
2.	**Change compared to preferred model**	**Manyara data**	3568	1588	892	1572	252
3.		**Power 0.9**	3048	1356	764	2104	340
4.		**BL with corr 0.04[a]**	2272	1012	568	1568	252
5.		**BL with corr 0.2[a]**	2184	972	548	1508	244

Note : [a] These numbers are based on using ANCOVA for analysis.

Source : Author's compilation

Overall, these power calculations show that more realistic sample sizes mean that detectable effect sizes are 15 percent or larger. Detecting an achieved maize yield increase of 133.7 kg per acre (a 15 percent increase) in Babati under this intervention requires 1,012 farmers in the preferred base case and 972 farmers if we can assume an inter-temporal correlation that is larger than what I found in the ERHS data. Therefore, for the present intervention, 1,000 is the recommended sample size.

Data

Each work package collects administrative data which sometimes includes identification (and some demographic) data on project beneficiaries. So far, however, these represent ad-hoc data capture exercises and are not sufficient for purposes of evaluation. To extend the information set, it is proposed to collect three rounds of survey data, a baseline and two follow-ups (one and three years after the intervention). Each round will include two types of questionnaires, a household survey and a community survey, in combination known as the Africa RISING Baseline Evaluation Survey (ARBES). Questionnaires

for both surveys have already been developed.[8] The evaluation will therefore use a new dataset for Tanzania. The specific purpose of the ARBES is to provide data for evaluation of Africa RISING activities in the East and Southern Africa sites. In Babati District a baseline of at least 800 households in six villages is planned but, for this evaluation, power calculations suggest this may have to be increased to 1,000 households (see 'Power Calculations' above). The community survey will collect data in each sub-village.

Table 2. Data sources for evaluation of Africa RISING in Babati

Item	*Observations*	*Sources*
ARBES (household): baseline 2013 and follow-up 2014 and / or 2016	At least 800 agricultural smallholder households	AR / IFPRI (either in-house or via external recruitment of a data-collection firm under IFPRI oversight)
ARBES (community): baseline 2013 and follow-up	Up to 29 communities (hamlets) in six villages	Same as above
AR / Babati program information (administrative records, monitoring data): 2011-2016		Work package research teams, key informant interviews
Agricultural and economic data: 2002-2016		Government of Tanzania, international agencies

Source: Author's compilation

In addition to a household roster that captures individual-level demographic information (relationship to head, sex, age, education, and marital status), the household survey includes individual- and household-level modules (as warranted) on consumption (short-term food expenditure, medium-term non-food expenditure, and annual expenditure on clothes, education, health, home improvement, and major family events), health (financing, recent illnesses), anthropometrics (for children under five years old and women of reproductive age), labor (on- and off-farm activity, wage versus non-wage), food security and subjective welfare, assets and housing characteristics (roof, wall, and floor material, ownership of appliances and farm equipment), livestock (feed type, source, cost), and crop operations (size, tenure, crop type, input types, production, sales, storage). The community survey includes a roster of key informants and several modules that cover access to services, land use, access to extension, other projects operating in the community, and local prices.

The baseline and follow-up surveys will be restricted to six villages in Babati District: Bashnet, Long, Loto, Magwanjwa, Sabilo, and Seloto (See 'Design' above). They will cover all households that are part of the evaluation sample.

Table 3 presents a template for summary statistics to be described once the baseline has been completed using a few illustrative indicators. These will be done at the household level and the individual level (working-age adults, children under five). Information for empty cells of this table (and for other indicators such as demographic characteristics, housing amenities, nutrition, and labor market status) will be provided by the baseline survey.

[8] Questionnaires are available upon request.

Table 3. Baseline (2013) summary statistics of the AR-in-Babati evaluation sample*

	Household: Maize yield					Individual: Percent of working-age adults employed in agriculture					Individual: Proportion of under-five children underweight				
Village	Mean	Sd	Min	Max	Obs	Mean	Sd	Min	Max	Obs	Mean	Sd	Min	Max	Obs
Bashnet															
Long															
Loto															
Magwanjwa															
Sabilo															
Seloto															
Total					**1,000**										

Note: * Empty cells to be filled in after baseline data collection; [a] Sd – standard deviation, Min – minimum, Max – maximum, Obs – number of observations. These outcomes reflect the fact that key outcomes will be measured at both household- and individual-level.

Source: Author's compilation

Threats to evaluation design

A few threats to internal validity remain and must be taken into account. For the input provision intervention, it is important to ensure that the right amounts of improved seeds or fertilizer are given to farmers. Farmers given more than required (for their plot size) may choose to give or sell away any excess, potentially leading to spillover and contamination of the control group of farmers. In addition, the first follow-up survey should measure input sharing and input sales.

For the training intervention, due to the web of individual-partner-led components under Africa RISING, the evaluation design faces the challenge to ensure that project implementation follows the randomization scheme outlined here. Attendance at the training sessions therefore needs to be carefully monitored. Furthermore, there might be concerns about knowledge spillovers, that is, that farmers who receive the training intervention share what they have learned with their neighbors who did not receive the training. However, studies on the impact of farmer field schools have found no evidence that knowledge acquired by treatment farmers is diffused to other farmers (Feder et al. (2004b)).

For both interventions, even if their content (input and / or training) itself is not shared with control farmers, social learning may still pose a threat to the internal validity of the proposed evaluation. As has been shown in the literature on social learning and technology adoption, farmers may learn from the experience of others with a new crop and condition their own growing decisions on this (Foster and Rosenzweig (1995)), Munshi (2004), Bandiera and Rasul (2006), Conley and Udry (2010)). The first follow-up survey should additionally measure the source of any new technologies applied during the just-completed cropping season.

Conclusion

Africa RISING presents a challenge to evaluation from many perspectives, not least because of its complicated network of research activities and varied implementation schedule. This note proposed a prospective evaluation design for Africa RISING activities in a district of northern Tanzania and an empirical method based on randomizing farming households in six villages into groups receiving inputs, training, or a combination of the two treatments.

By no means is this design meant to exhaust all the evaluation possibilities likely to be revealed as the program reaches full operational status. For example, using these early results, additional experiments building on what we will learn from the proposed evaluation are possible, such that, if the direct provision of inputs is shown to work, a further step could be to test different subsidy levels. If instead training is shown to be most effective, testing the use of text messages for updates and reminders about the training content could be considered, an inexpensive but potentially high-impact intervention.

In summary, the proposed design does present an early-stage opportunity for researchers to assess whether farmers drawn to similarly configured and executed programs (and their families) can benefit both at plot and household level, and may point the way toward future research. From a policy perspective, this new evidence will also be invaluable in determining whether AR trials merit further support and possible scaling up.

References

Annan, Jeannie, Charles Dixon, Rachel Glennerster, Frances Kimmins, and Tavneet Suri. n.d. *Promoting adoption of new rice varieties: Addressing the costs of early adoption in Sierra Leone*. Cambridge, MA: Abdul Latif Jameel Poverty Action Lab (Ongoing study).

AR. 2011. *Program framework 2012-2016*. (Mimeo).

AR. n.d. *Research in sustainable intensification in the sub-humid maize-based cropping systems of Babati: Description of the intended work plan*. (Mimeo).

AR. n.d. *Intensification of maize-legume based systems in the semi-arid areas of Tanzania (Kongwa and Kiteto districts) to increase farm productivity and improves farming natural resource base*. (Mimeo).

AR. 2013. *Babati and Kongwa-Kiteto interim progress reports*. (Mimeo).

Bandiera, O. and I. Rasul. 2006. Social networks and technology adoption in northern Mozambique. *The Economic Journal* 11: 869–902.

Benin, S., E. Nkonya, G. Okecho, J. Randriamamonjy, E. Kato, G. Lubade, M. Kyotalimye, and F. Byekwaso. 2008. *Impact evaluation and returns to investment of the national agricultural advisory services (NAADS) program of Uganda*. Discussion Paper 724. Washington, DC: International Food Policy Research Institute.

Bolwig, S., P. Gibbon, and S. Jones. 2009. The economics of smallholder organic contract farming in tropical Africa. *World Development* 37(6): 1094–1104.

Carter, M. and D. Yang. n.d. *Savings, subsidies and sustainable food security in Mozambique*. New Haven, CT: Innovations for Poverty Action (Ongoing study) (http://www.poverty-action.org/project/0384).

Conley, T. G. and C. R. Udry. 2010. Learning about a new technology: Pineapple in Ghana. *American Economic Review* 100 (1): 35-69.

Dercon, S., D.O. Gilligan, J. Hoddinott, and T. Woldehanna. 2008. *The impact of agricultural extension and roads on poverty and consumption growth in fifteen Ethiopian villages*. Discussion Paper 840, Washington, DC: International Food Policy Research Institute (IFPRI).

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *American Economic Review* 101: 2350–2390.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2008. How high are rates of return to fertilizer? Evidence from field experiments in Kenya. *American Economic Review* 98(2): 482–88.

Essama-Nsaah, B. 2013. Coping with the attribution problem in program evaluation. Special Series on the Nuts and Bolts of M&E Systems, number 26. Washington, DC: World Bank.

Farley, Katherine, Sarah Lucas, Jack Molyneaux, Kristin Penn, and Emily Hogue. 2012. *Impact evaluations of agriculture projects*. Millenium Challenge Corporation Principles into Practice series. Washington, DC: Millenium Challenge Corporation.

Feder, G., R. Murgai, and J.B. Quizon. 2004a. Sending farmers back to school: The impact of farmer field schools in Indonesia. *Review of Agricultural Economics* 26 (1): 45–62.

Feder, Gershon, Rinku Murgai, and Jaime Quizon. 2004b. The acquisition and diffusion of knowledge: The case of pest management training in farmer field schools, Indonesia. *Journal of Agricultural Economics* 55(2): 221-243.

Foster, A., and M. Rosenzweig. 1995. Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy* 103(6): 1176-1209.

FSG. 2010. *Smallholder Farmer Development: International Donor Funding Trends. A Trend Analysis of the Corporate Leadership Coalition for Smallholder Farmer Livelihoods*. Boston: Foundation Strategy Group.

(http://www.fsg.org/Portals/0/Uploads/Documents/PDF/Smallholder_Development_Donor_Funding_Trends.pdf?cpgn=WP%20DL%20-%20Smallholder%20Donor%20%28FULL%29)

Gertler, Paul, Sebastian Martinez, Patrick Premand, Laura B. Rollins, Christel M.J. Vermeersch. 2011. *Impact evaluation in practice*. Washington, DC: The World Bank.

(http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf) or (http://www.worldbank.org/pdt)

Godtland, E., E. Sadoulet, A. De Janvry, R. Murgai, and O. Ortiz. 2004. The impact of farmer field schools on knowledge and productivity: A study of potato farmers in the Peruvian Andes.‖ *Economic Development and Cultural Change* 53 (1): 63–92.

Gollin, Douglas. 2006. *Impacts of international research on intertemporal yield stability in wheat and maize: An economic assessment*. Mexico, D.F.: International Maize and Wheat Improvement Center (CIMMYT).

IDB. 2010. *Development effectiveness overview: Special topic – Assessing the effectiveness of agricultural interventions*. Washington, DC: Inter-American Development Bank.

IMF. 2012. Tanzania: Poverty reduction strategy paper—joint staff advisory note. *Country Report No. 12/118*. Washington, DC: International Monetary Fund.

IMF. 2013. United Republic of Tanzania: Sixth review under the policy support instrument, second review under the standby credit facility arrangement, and request for modification of performance criteria—staff report; Press release on the executive board discussion; and Statement by the executive director for Tanzania. *Country Report No.* 13/166. Washington, DC: International Monetary Fund.

McKenzie, David. 2012. Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics* 99(2): 210-221.

Munshi, K. 2004. Social learning in a heterogeneous population: Technology diffusion in the Indian green revolution. *Journal of Development Economics* 73: 185– 213.

NBS. 2013. *2012 Population and housing census: Population distribution by administrative areas*. Dar es Salaam: National Bureau of Statistics.

NBS. 2012. *National Sample Census of Agriculture 2007 / 2008: Volume Vu: Regional Report: Manyara Region*. Dar es Salaam: National Bureau of Statistics.

NBS. 2010. *National Sample Census of Agriculture 2007 / 2008: Preliminary Report*. Dar es Salaam: National Bureau of Statistics.

Owens, T., J. Hoddinott, and B. Kinsey. 2003. The impact of agricultural extension on farm production in resettlement areas of Zimbabwe. *Economic Development and Cultural Change* 51 (2): 337–357.

StataCorp. 2009. *Stata statistical software: Release 11*. College Station, TX: Stata Corporation.

World Bank. 20. *Impact evaluations in agriculture: An assessment of the evidence*. Washington, DC: The World Bank.